

DIAGNOSTIC AND SCREENING METHOD

FIELD OF THE INVENTION

The present invention relates to a novel diagnostic method for the detection of infertility in males. In particular, the present invention relates to a diagnostic method for detecting the presence or absence of a mutation or mutations in the *POLG* gene in a biological sample. The invention relates also to the use of a mutant *POLG* gene in the detection of infertility in males and in the screening of human populations for the presence of such mutation or mutations as a predictive test for male infertility. The present invention further relates to the use of the *POLG* gene as an indicator of other pathological conditions associated with or related to male infertility, including those manifesting in women.

BACKGROUND OF THE INVENTION

Fertility problems have manifested increasingly in western countries due to a variety of social causes. For example, young couples delay their decision to establish a family for reasons of education and career, and tend to plan for a smaller family size due to the increasing priority placed upon quality of life. Environmental and lifestyle factors may also play a part in revealing an underlying subfertility due primarily to other causes. For many couples, prolonged attempts to become pregnant end in failure, resulting in their seeking assistance from an infertility clinic. This decision is often taken long past the time when they initially would have hoped to start a family. People are understandably reluctant to take a step that may seem humiliating. Such delays, together with the attendant stress, anguish and shame, can have a profoundly destructive effect on relationships and on quality of life generally.

Infertility can result from a great variety of causes, including anatomical, developmental, infectious and toxicological factors. The majority of cases can be attributed wholly or predominantly to one or other partner, with roughly equal numbers of cases of male and female factor infertility. In both instances, however, the primary cause is almost certainly genetic. Estimates of the true population prevalence of infertility vary, but are generally accepted to be in the range of 2 to 5%. This is, therefore, one of the commonest human genetic disorders.

Male infertility can manifest as reduced quantity (oligozoospermia) or total lack (azoospermia) of sperm, reduced quantity of motile sperm

(asthenozoospermia) or morphologically abnormal sperm (teratozoospermia). These categories are not exclusive, since many infertile males appear to fall into more than one such class. Microscopic, cytological and biochemical methods are presently available for the determination of these defects. However, actual measurements can be highly variable, even for a single individual between samplings. Frequently, moreover, no actual cause for male infertility can be found. This can cause additional costs arising from the need for further clinical examinations. Selecting an appropriate method of assisted reproduction is often a haphazard and costly matter, with no guarantee of eventual success. All these difficulties and delays can place severe, additional strains on relationships. Additional means for the early detection or prediction of infertility, in particular male infertility are thus needed. Early detection of male infertility would save both money and human stress and would allow appropriate counselling or assisted reproduction to be offered to a significant population of individuals who would otherwise only discover their problem after it has caused severe damage to their quality of life.

Though it has been long suspected that male infertility in most cases is a genetic disorder, and furthermore that it is genetically heterogeneous (i.e. that many different genes are the underlying cause of the disorder in different individuals), little information has emerged relating specific genetic defects and male infertility. The two known genetic causes of male infertility are the 'cystic fibrosis' mutations in the *CFTR* gene, part of a clearly recognizable clinical syndrome affecting approximately one person in 1500 [see, e.g. Lissens, W. and Liebaers, I. *Baillieres Clinical Obstetrics and Gynaecology* 11 (1997) 797-817; Schnedl, W. et al., *Wiener Klinische Wochenschrift* 103 (1991) 29-33] and a deletion on the Y chromosome, that leads to complete azoospermia, and is found in approximately 1 to 2% of infertile males in this category [Elliott, D.J. and Cooke, H.J. *BioEssays* 19 (1997) 801-809]. The supposed genetic cause of male infertility in the vast majority of cases has thus far remained unknown.

Since spermatozoa are heavily dependent on respiratory energy for motility, impaired energy metabolism, whether in mature spermatozoa or at earlier stages of male germ cell differentiation, is a long hypothesized mechanism contributing to infertility. Defects in mitochondrial function [Johns, J.C.S. et al., *Nat Med* 3 (1997) 124-125], possibly associated with mtDNA lesions [Kao, S.H. et al., *Mol Hum Reproduction* 4 (1998) 657-666; Lestienne P. et al.,

Mol Hum Reproduction 3 (1997) 811-814], have been reported in sperm samples from infertile males, and sperm motility appears to be correlated with mitochondrial respiratory activity [Ruiz Pesini, E. et al., Clin Chem 44 (1998) 1616-1620] and membrane potential [Troiano, L. et al., Exp Cell Res 241 5 (1998) 384-393]. However, no relationship between a specific mitochondrial gene and male infertility has been disclosed or even suggested.

SHORT DESCRIPTION OF THE INVENTION

It has now, surprisingly, been discovered that the frequency of a mutant genotype in the *POLG* gene encoding the catalytic subunit of mitochondrial DNA polymerase (DNA polymerase γ) is significantly increased in groups of infertile males. Specifically, the mutant genotype has been located in the area of the trinucleotide (CAG) microsatellite repeat of the *POLG* gene within the N-terminal region of the coding sequence. While the normal *POLG* gene contains 10 consecutive, glutamine encoding CAG codons, followed by a 10 single CAA and two further CAGs, at least one allele with an altered CAG repeat-length in the *POLG* gene is found in a large population of infertile men. 15

The *POLG* mutations described in the present application represent the first significant step in efforts of elucidating the genetic nature of male infertility and provide novel means of diagnosis thereof.

20 An object of the invention is thus to provide a non-invasive, non-intrusive diagnostic method that is useful in identifying, detecting and characterizing male infertility in the large fraction of cases where its causes remain unsolved.

25 A second object is to provide a simple screening test of strong predictive value, in order to identify in advance those individuals who will require reproductive counselling and assistance. Such information will greatly enhance the expected quality of life of those who suffer from this disorder.

The present invention relates to a new method for the diagnosis of male infertility by detecting the presence or absence of a mutation or mutations 30 in the *POLG* gene encoding the catalytic subunit of mitochondrial DNA polymerase in a biological sample.

The present invention also relates to the use of a mutant form of the *POLG* gene encoding the catalytic subunit of mitochondrial DNA polymerase for the diagnosis of male infertility.

35 The present invention also relates to the use of a mutant form of the *POLG* gene encoding the catalytic subunit of mitochondrial DNA polymerase

as a diagnostic agent.

The present invention further relates to diagnostic kits containing suitable reagents to detect a mutant form of the *POLG* gene or the normal *POLG* gene.

5 The present invention further relates to the use of the *POLG* gene as an indicator of other pathological conditions associated with or related to male infertility, including those manifesting in women.

DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates the genotype data of Table 2 as a bar chart.

10 DETAILED DESCRIPTION OF THE INVENTION

Microsatellite sequences (short sequences consisting of consecutive runs of a simple repeated element, such as a di- or trinucleotide) are well established sources of genetic polymorphism, exhibiting significant levels of length variation in the many genetic loci where they are found. This reflects the 15 fact that they are inherently less stable than other sequences during, for example, DNA replication or repair, where slippage mis-pairing events may increase or decrease the number of re-iterated units. Many such sequences are found within the human genome, and are usually located outside of coding regions, where they have little consequence for phenotype. In some instances, 20 however, they are located within or adjacent to coding DNA. In these cases, large expansions of the number of repeat units can have a profound effect, leading to loss or gain of genetic function and a resulting pathological phenotype. Instability at trinucleotide repeats is already known to be associated with various human disorders, including, for example, Huntington's disease (HD), 25 myotonic dystrophy (DM), and several forms of spinocerebellar atrophy (e.g. SCA1). Other, more subtle variations at trinucleotide microsatellites can also affect phenotype, and the association described herein, between male infertility and variation in the CAG trinucleotide tract of the *POLG* gene, encoding the catalytic subunit of mitochondrial DNA polymerase, falls into such a category.

30 Rovio, A. et al. studied the frequency of different repeat-length alleles of the *POLG* gene in control populations, as well as in groups of patients suffering from recognized mitochondrial disorders (Eur J Hum Genet, 1999 in press). The predominant allele of 10 CAG repeats was found at a very similar frequency (approximately 88%) in both Finnish and ethnically mixed population samples, with homozygosity close to the equilibrium prediction of 80%.

Other alleles of between 5 and 13 repeat units were detected, but no larger, expanded alleles were found. Patients with a variety of molecular lesions in mtDNA showed no differences in *POLG* trinucleotide repeat-length distribution from controls. The authors concluded that their findings rule out *POLG* repeat 5 expansion as a common pathogenic mechanism in disorders characterized by mitochondrial genome instability.

The present invention is based on studies that were designed to test a hypothesis that this gene, encoding a major component of the machinery of mitochondrial DNA maintenance, is involved in infertility, via impaired 10 energy metabolism of spermatozoa. The *POLG* gene encodes a protein of approximately 130 kDa and has been mapped to the region of chromosome 15q24-15q26 [Walker, R.L. et al., *Genomics* 40 (1997) 376-378; Ropp, P.A. and Copeland, W.C., *Genomics* 36 (1996) 449-458; Zullo, S.J. et al., *Cytogenet Cell Genet* 78 (1997) 281-284].

15 For PCR analysis, nested primers flanking the *POLG* CAG repeat were designed, based on a region of the coding sequence of *POLG* (see sequence id. no. 1). The primers amplified the predicted fragments from genomic DNA, as verified by direct sequencing. Testing against the monochromosomal interspecies hybrid and Genebridge 4 radiation hybrid panels [Gyapay, G. et 20 al, *Hum Mol Genet* 5 (1996) 339-346] confirmed that they detected only the expected gene sequence from chromosome 15q, and not a pseudogene. The 5' primer (also designated mip51) was also found to generate the predicted products, using other, more downstream 3' primers from the same exon of the *POLG* gene. Using these primers and sperm DNA from individuals in whom 25 azoospermia was excluded, the PCR analysis showed that 9 out of 99 male infertiles were homozygous for absence of the normal, 10 repeat-unit allele. Thus approximately 10% of infertile males, whose infertility is not clearly attributable to known causes, and excluding cases of azoospermia, completely lacked the common *POLG* allele, having 10 consecutive CAG codons. Instead 30 they possessed two alleles with a different number of CAG repeats. No instances of this genotype were found in sperm DNA from 98 fertile males. The association is highly significant based on two different statistical tests (the z-test for percentages, as well as Poisson probabilities). The homozygous mutant genotype was found in only 2 of over 250 healthy controls collected with- 35 out regard to gender or fertility.

Male infertiles also showed a higher frequency of heterozygosity

than found amongst fertile males or in the general population. Some 35% of infertile males, compared with only about 20% in the fertile or unselected population, were heterozygous at the *POLG* CAG repeat, i.e. possessed one copy of the wild-type or normal allele with 10 repeats, and one other copy.

5 These figures strongly indicate that many of the 35% of infertiles who are heterozygotes are, in fact, compound heterozygotes, i.e. that the allele of the *POLG* gene that is apparently wild-type or normal at the CAG repeat in fact carries a mutation elsewhere in the gene, such that both copies are defective, at least as regards the function that *POLG* performs in spermatogenesis that is
10 required for fertility.

PCR studies of *POLG* genotype frequencies in blood DNA from various categories of patients indicated that a mutation at this locus is not associated with azoospermia or with severe oligozoospermia, but is found in individuals with a range of sperm quality defects.

15 Also fluorescent PCR using custom primers was used to genotype all individuals in the study for *POLG* CAG repeat length, as summarized in Table 2. Semen DNA from a minority (approx. 15%) of male infertiles yielded no clear signal and these were excluded from consideration. Approximately 10% of individuals lacked the normal allele. By contrast, this normal allele was present in all 98 fertile males studied. The normal allele was also absent in only one out of 118 healthy male controls, which had not been selected on the basis of fertility, in addition to the 250 controls studied previously without regard to gender or fertility.

Given that the combined frequencies of mutant alleles (i.e. those
25 having any number of repeats other than 10) revealed in the population survey of Rovio, A. et al. (supra) was 0.12, the homozygous mutant genotype is expected to be found in approximately 1.5% of randomly selected individuals, based on standard Hardy-Weinberg predictions (see Table 2). The above data indicate that it was found in the general population slightly below expectation
30 (0.8%, i.e. 2 out of 252), although this deviation is not statistically significant. By contrast, finding the homozygous mutant genotype by chance in 9 out of 99 individuals is exceedingly unlikely, assuming that the previously measured allele frequencies are representative. The association is significant ($p < 0.01$) based on a z-test for percentages (comparing fertile and infertile groups), or
35 using Poisson statistics (assuming a mean population frequency of the homozygous mutant genotype of 1.5%).

WO 00/50629

PCT/FI00/00140

Many different alleles and combinations thereof were found in individuals lacking the common allele (Table 1). This indicates that it is the absence of the normal allele, rather than the presence of a particular alternate allele, that is associated with the phenotype. All alleles found in infertiles were 5 also found in controls. The frequencies, relative to one another, of the various mutant alleles found amongst infertiles were also similar to those found in the general population. The different profiles of *POLG* genotypes found in sperm DNA from fertile and infertile males, compared with controls, is illustrated in Figure 1. 35 out of 99 (i.e. 35%) of infertiles were heterozygotes, carrying just 10 one copy of the normal allele.

This is well above expectation (21%), based on the measured allele frequencies in the overall population, or compared with the fertile control group (18%). Importantly, some of these could represent compound heterozygotes, with a second mutation mapping elsewhere in the gene than within the CAG 15 repeat of the coding region of the gene.

In order to establish the type of infertility associated with *POLG* genotype, blood DNA was studied from two larger groups of patients, who included a high proportion of azoospermic individuals and those with severe oligozoospermia (sperm concentration $< 5 \times 10^6 / \text{ml}$), as well as individuals with 20 only a mild abnormality of sperm concentration, motility and/or morphology. Sperm DNA was also analyzed from a further set of patients, for whom extensive phenotypic data were available, and which consisted of cases of 'pure asthenozoospermia', with sperm concentrations well into the normal range (generally $> 100 \times 10^6 / \text{ml}$).

25 Cases lacking clear results of sperm analysis were excluded, as were those where clinical or karyotypic examination clearly indicated an established hormonal, developmental or traumatic etiology of their infertility. The azoospermic and severe oligozoospermic patients were considered separately from those having sperm concentrations of $5 \times 10^6 / \text{ml}$ or greater.

30 Out of 62 azoospermic and 73 severe oligozoospermic individuals, no cases of homozygous absence of the normal allele were detected. The frequency of heterozygotes in these groups (30%) was, however, slightly above the population average. By contrast, 8 out of the 113 with infertility of unknown etiology, and having measured sperm concentrations of $5 \times 10^6 / \text{ml}$ or greater, 35 lacked the normal allele.

Individuals homozygous for the absence of the normal allele were

found in every category of sperm quality defect (sperm number, motility or morphology, or any combination). Most, however, were at or below at least two of the thresholds commonly used to define oligozoospermia (20×10^6 sperm/ml), asthenozoospermia (50% motile sperm) and teratozoospermia

- 5 10% morphologically normal spermatozoa). However, no mutant genotype was found in semen DNAs from the 56 patients with 'pure asthenozoospermia'. It is concluded that the mutant *POLG* genotype is associated with a general impairment in sperm quality and/or number, but neither with the complete or almost complete absence of spermatozoa, nor with a motility defect as
- 10 such. The simplest interpretation is that cell division or differentiation is impaired, leading to a low production of normal spermatozoa.

The above data clearly indicate that *POLG* alleles other than the normal one of 10 repeat units are deleterious to sperm function or differentiation, but have little or no effect on other tissues. Polyglutamine tracts are

- 15 commonly regarded as interfaces for protein-protein interactions, hence it is possible that a sperm-specific protein interacts with this region of the polypeptide. Other possibilities are that repeat-length variants are inherently unstable, poorly targeted to mitochondria, or enzymatically defective, all of which are testable. The many rounds of cell division during spermatogenesis, plus the
- 20 necessity, for the creation of functional spermatozoa, of maintaining a genetically fit mitochondrial genome, argue that a sub-optimal mtDNA polymerase, whether by virtue of impaired fidelity, processivity, or some other defect, could result in the accumulation of mtDNA mutations and failure to complete differentiation.

- 25 Male infertility is certainly a genetically heterogeneous disorder [Sokol, R.Z., Curr Opinion Obst Gynaecol 7 (1995) 177-181]. The data of the present invention indicate that where there is no other obvious etiology, approximately 10% and perhaps as many as 20% of cases are associated with a mutant genotype at the *POLG* locus, excluding those where sperm count is
- 30 negligible.

The measured frequency of mutant alleles in the population (approx. 0.12) means that approximately 1.5% of all males will lack the normal allele, and are predicted to be infertile or at least subfertile, based on these findings.

- 35 According to the diagnostic method of the present invention, the presence or absence of a mutation or mutations in the *POLG* gene can be

detected from a biological sample by any known detection method suitable for the purpose. Thus, the presence or absence of a mutation or mutations in the *POLG* gene can be detected from a biological sample by any known method for detecting gene copy number or expression, i.e. methods based on detecting the copy number of the gene (or DNA) and/or those based on detecting the gene expression products (mRNA or protein). Such methods are easily recognized by those skilled in the art and include the use of the polymerase chain reaction (PCR) or other thermal cycler-based DNA synthetic techniques in the presence of appropriate oligonucleotide primers specific for the *POLG* gene, 5 molecular cloning in a plasmid or other suitable vector, detection of length variants in a DNA sample by agarose or polyacrylamide gel electrophoresis or any analogous technique, gel or capillary electrophoresis with automated detection and analysis of products tagged with a fluorescent or other label incorporated into the DNA, DNA sequence determination by any technique, hetero- 10 duplex-based and associated methods for detecting base mismatches or length variants, mass spectroscopy, chromatographic and other separation methods, Western analyses, immunohistochemistry, and other immunoassays, and any technique suitable for the detection and functional characterization of nucleic acid and protein which may be applied in mutational analysis. Prefer- 15 able methods are those suitable for use in routine clinical laboratories, or which may eventually be developed in self-test kits.

In the diagnostic method of the invention, the biological sample can be any suitable tissue sample or body fluid, such as buccal smear (mouthwash), hair roots, semen, seminal plasma, whole blood, serum or 20 plasma, or cultured cells derived from any biopsy. The biological sample can be, if necessary, pretreated in a suitable manner known to those skilled in the art, for the purposes of extraction of DNA in a form suitable for analysis.

The diagnostic kit of the present invention comprises reagents necessary for the detection of a mutation or mutations in the *POLG* gene. These 25 reagents include specific antibodies, preferably monoclonal antibodies, capable of identifying the *POLG* gene or its gene products, other antibodies, primers, markers and/or standards that are needed for visualization or quantification as well as buffers, diluents, washing solutions and the like, commonly contained in a commercial reagent kit. Alternatively, the diagnostic kit of the 30 present invention may comprise portions of the *POLG* gene or its functional variant or fragment, or other nucleic acid-derivative related to it, together with 35

suitable reagents, such as those listed above, needed for the detection of the mutation or mutations in the *POLG* gene.

The present invention provides an additional means for detecting and analyzing male infertility, and a means for predicting it, when carried out 5 on adults, children, in infancy, *in utero* or during embryonic development.

The invention is now elucidated by the following non-limiting examples. The patients and controls used in the first example are as follows. Volunteers for the study, as authorized by the local ethical committee, were recruited from amongst male infertiles attending an infertility clinic in Oxford.

10 Semen samples obtained by masturbation were stored frozen in liquid nitrogen for further study. Azoospermic individuals (approx. 20% of the total) were excluded, as were any for whom a clear, established etiology (hormonal, developmental, karyotypic or traumatic) could be found to account for their infertility. Semen was also collected from sperm donors, for use as fertile controls in the 15 study. Semen or seminal plasma was also provided from similarly defined infertile males in Tampere, Finland, and from infertiles in Australia and Taiwan. Blood samples were provided from fertile and infertile males, as well as from unselected control males in Düsseldorf, Germany, and from male infertiles in Edinburgh, Scotland. Blood samples were also provided by otherwise healthy 20 volunteers amongst laboratory personnel and associates, who were not screened for fertility.

Example 1

Genotyping at the *POLG* CAG repeat as a diagnostic test for male infertility

25 DNA was prepared from fresh or frozen sperm or seminal plasma as follows. 400 μ l samples were micro-centrifuged at 13,000 rpm for 5 min and the pellets washed twice with phosphate buffered saline, pH 7.4 (PBS). Crude DNA was then extracted from the washed pellets by incubating with proteinase K as described by Reid, F.M. *et al.* [Hum Mutation 3 (1994) 243-247], followed 30 by inactivation of the enzyme at 92°C for 10 minutes. Blood DNA was isolated by the same method, as described by Reid, F.M. *et al.*, *supra*. DNA can be obtained from other biological samples using essentially similar methods which are well known to those skilled in the art. Prior to all PCR analyses, the DNA extracts were microcentrifuged for 15 sec and 0.5 μ l of clear lysate used as 35 template. In any cases where signals were low or uninterpretable, PCR was repeated using 0.5 μ l of the template DNA solution diluted to various extents in

order to reduce the concentration of contaminants that may inhibit DNA synthesis. More than 0.5 μ l of template DNA solution may also be used, if sample concentration is judged to be too low.

Fluorescent PCR analysis of *POLG* genotype was carried out as follows. Nested oligonucleotides corresponding to regions of the N-terminal coding sequence of *POLG*, as indicated in the accompanying list of sequences, were purchased from Life Technologies (Paisley, Scotland) and DNA Technology (Aarhus, Denmark). One of the primers, mip51 (5' CCAGCTCCGTCCCCGCGTCCGACC 3'; sequence id. no. 2), was 5' prelabelled with the fluorescent dye ROX (PerkinElmer) by the manufacturer, but other dyes of similar application can also be used.

Fluorescent PCR reactions used, at 0.2 mM each, ROX-labeled primer mip51 plus one downstream primer, unlabelled, either mip31 (5' GCTGCCCGCCCTCCGAGGGATAGCAC 3'; sequence id. no. 3; generating a 126 bp PCR product from the wild-type allele, when used in combination with primer mip51), mip32 (5'CTCTCGAGAGCATCTGGATGTCCAATC 3'; sequence id. no. 4; generating a 165 bp PCR product from the wild-type allele, when used in combination with primer mip51) or mip33 (5'CTCGTGCAGCCCTCTGAGAGCAT 3'; sequence id. no. 5; generating a 176 bp PCR product from the wild-type allele, when used in combination with primer mip51). Reactions were carried out in 12.5 μ l at 200 μ M dNTPs (Pharmacia Biotech, Sweden), plus an appropriate quantity of a suitable thermostable DNA polymerase (0.15 units of Dynazyme, from Finnzymes, Espoo, Finland) in the manufacturer's buffer. Reactions comprised 30 cycles of denaturation for 1 min at 95°C, annealing for 45 sec at 62°C, and extension for 1 min at 72°C (a 5 min extension in final cycle). Products were diluted 1:10 in water, and samples containing 1 μ l of diluted PCR product, 12 μ l of deionized formamide and labelled DNA fragment size markers in the appropriate size range were added [e.g. 0.15 μ l of Tamra Genescan350 DNA size standards (Perkin Elmer)], and samples were analyzed by capillary electrophoresis on a suitable electrophoresis/fluorescent analysis instrument, e.g. ABI 310 Genetic Analyzer (PerkinElmer), using the manufacturer's data collection and analysis software.

Samples giving ambiguous or low signals, or those giving a result indicative of a mutant genotype, were re-evaluated using a second downstream primer. In all analyses, a reaction using a template DNA sample previ-

ously demonstrated to contain only the wild-type *POLG* allele, as well as a reaction run without any added DNA, were run alongside as positive and negative controls, respectively.

5 The results showing the mutant *POLG* alleles of male infertiles are shown in Table 1.

Table 1
Mutant *POLG* alleles of infertile males

<i>POLG</i> repeat-length genotype	Number
11/11	8
11/12	5
11/9	3
9/12	2
11/13	1
9/9	1
8/11	1
8/12	1
6/12	1
Total	23

10

The PCR amplification and fluorescent analysis described enabled samples to be categorized as follows:

Class I – WILD-TYPE (or normal) HOMOZYGOTES: only a single allele was detected, as a clean peak of a size consistent with the presence of 15 10 consecutive copies of the CAG repeat.

Class II – MUTANT 'HOMOZYGOTES': one or two alleles were detected, both corresponding with length variants other than the wild-type allele of 10 CAG repeats. These individuals are compound heterozygotes at the *POLG* CAG repeat. However, since such individuals are homozygous for loss 20 of the wild-type allele, they are placed in a different category from 'true' het-

erozygotes as defined below (Class III).

Class III – HETEROZYGOTES: two length alleles were detected, one corresponding with the wild-type allele (10 CAG repeats) and the other a different number.

5 Data for such an analysis are illustrated in Figure 1 and Tables 2 and 3, below. Controls were the combined sets collected previously from Finland, being a mixture of Finns plus other ethnic groups of both sexes, plus those from Germany (males only, no selection based on fertility). Infertiles and fertiles indicated in Table 2 each represent pooled groups of individuals from
10 Finland and England.

Table 2

POLG genotype frequencies detected in various fertile and infertile males (sperm), plus unselected controls (blood)

Patients (COUNTRY)	Class I: wild-type homozygotes	Class II: mutant 'homozygotes'	Class III: het- erozygotes
Infertile males (FIN/ENG)	55 (56%)	9 (9%)	35 (35%)
Fertile Males (FIN/ENG)	80 (82%)	0 (0%)	18 (18%)
Total controls (FIN/ENG/D)	195 (77%)	2 (1%)	55 (22%)

Table 3
Frequency of 'mutant homozygotes' (Class II) in various groups of infertile males

Patients (COUNTRY, DNA SOURCE)	Total analysed	Class II: mutant 'homozygotes'
Azoospermia or severe oligozoospermia (sperm concentration $< 5 \times 10^6/\text{ml}$) (D/SCO, blood)	135	0 (0%)
Pure asthenozoospermia (< 50% motile, normal sperm concentrations) (Taiwan, sperm)	55	0 (0%)
Combined sperm quality defect (2 or more of moderate oligozoospermia, as- thenozoospermia, and teratozoospermia) (D, blood)	113	8 (7%)

5

The analysis resolved the samples into the following groups in a clearly predictive manner, as follows.

(1) Genotype frequencies in the control group were very close to the equilibrium predictions based on overall population-based measurements
 10 of allele frequencies of 0.88 (wild-type, 10 CAG repeats) and 0.12 (all mutant alleles combined).

(2) No instances of 'mutant homozygotes' (Class II) were detected amongst fertile males.

(3) Approximately 9% of infertile males, excluding cases of azoo-
 15 spermia and severe oligospermia, fell into Class II.

(4) Heterozygotes (Class III) were found in all groups, but at a higher frequency in infertile than fertile males or controls.

The diagnostic and counselling implications of the test for *POLG* genotype that forms the subject of the present invention, and as proposed to
 20 be carried out, for example, in an infertility clinic, are as follows:

09-05-2000

+3100906399,1 - FI00/00140

R91

15

(1) A result falling into Class II is an indicator of a specific type of male factor infertility, associated with *POLG* dysfunction. This genotype is never found in fertile males. Other clinical or biochemical investigations of both partners can essentially be dispensed with as unnecessary. Even in the absence of sperm quality data, it indicates that intra-uterine insemination and probably IVF are likely to fail, and that recourse to ICSI (intra-cytoplasmic sperm injection) is advisable.

(2) A result falling into Class I does not exclude male factor infertility of other types, nor any other diagnosis.

10 (3) A result falling into Class III is ambiguous, and warrants further genetic investigation of the *POLG* locus, as illustrated below (example 2).

Example 2

Complete analysis of the *POLG* gene as a diagnostic test for male infertility

15 Results from the type of analysis illustrated in Example 1 are necessarily ambiguous in the case of individuals who prove to be heterozygous at the *POLG* gene CAG repeat (i.e. who possess one copy of the wild-type, 10-repeat allele, plus one copy of some other length variant). In these cases, further investigation is warranted, in order to establish whether they represent 20 cases of true heterozygotes (one fully functional, one mutant copy of the gene), as found amongst fertile males and controls, or whether they represent compound heterozygotes carrying one copy of the gene mutated within the *POLG* CAG repeat tract, and a second copy which has a pathological mutation elsewhere in the gene. The present invention thus covers not only the use of 25 genotyping at the CAG repeat, but also the complete sequencing of the *POLG* gene as a diagnostic procedure for the determination of male infertility associated with *POLG* gene dysfunction, in cases where genotyping at the CAG repeat is insufficient to permit an unambiguous molecular diagnosis. Because such cases are expected to constitute up to half of all individuals with *POLG*- 30 associated infertility, based on the observed frequencies of homo- and heterozygotes, i.e. given the excess number of heterozygotes amongst the infertiles, the further analysis illustrated in this Example will have an important diagnostic value, as regards the counselling and treatment to be offered.

35 DNA is prepared from the same sources and using the same methods as indicated under Example 1 (above). Long-extension PCR is carried out using a series of primer pairs flanking blocks of exons of the *POLG* gene cre-

ART 34 AMU

16

ating products that include the entire coding region of the gene, the 5' and 3' untranslated segments of the *POLG* mRNA, the proximal promoter of the gene, and all intron/exon boundaries. Oligonucleotide primers are based on the publicly available sequence of the *POLG* gene (deposited in the Genbank 5 database as accession number AC005317), and sourced from companies such as indicated under Example 1. In this specific example the relevant segments of the *POLG* gene are amplified initially as 3 large PCR products, although many other strategies are equivalent, based on other combinations of primers according to the database sequence. LX-PCR is carried out with a 10 standard reagent kit (e.g. Boehringer Mannheim Extend DNA Polymerase) under manufacturer's recommended conditions or using equivalent materials sourced elsewhere. PCR products are analysed by agarose gel electrophoresis and purified from agarose gels by standard, spin-column methods, e.g. using kits available from Qiagen or similar manufacturers. A total of 52 15 sequencing primers are also required (one in each direction for the 21 short exons and boundaries, two in each direction to cover the longer exon 23, and three in each direction to cover the longest exon, exon 2, in which the CAG repeat lies, presenting additional difficulties. The 20 bp sequences of these correspond with intronic sequences located approximately 30 bp up- and 20 downstream of the various intron-exon boundaries, based on the database-deposited gene sequence. In addition, internal exonic sequencing primers (one for each strand) are required for 20 bp regions in the middle of exon 23 and exon 2, as well as 20 bp on either side of the CAG repeat (one strand only, in each case, reading towards the repeat).

25 Fluorescent DNA sequencing is carried out using a standard dye-terminator reagent kit (e.g. PerkinElmer Big Dye) in thermal cycling reactions using the long PCR products as template and each sequencing primer as appropriate under the kit manufacturer's recommended conditions. Sequencing reaction products are analyzed by gel or capillary electrophoresis, on a suitable 30 electrophoresis/fluorescent analysis instrument, e.g. ABI 310 Genetic Analyzer (PerkinElmer), using the manufacturer's data collection and analysis software.

35 DNA sequences are carefully examined for evidence of 3 types of detectable heterozygosity, which manifest differently on the sequence traces, as follows. (a) A heterozygous point mutation manifests as a specific ambiguity localized to one nucleotide pair and detectable on both strands in a congruent

manner (i.e. the specific ambiguity on one strand is found as the exactly complementary ambiguity on the other). (b) A heterozygous deletion or insertion of one or more nucleotides in a localized region manifests as a break-point between a sequence that is readable (unambiguous) and a sequence that is totally unreadable (ambiguous at virtually every position), with the exact location and size of the deletion or insertion inferred by the positions at which ambiguity commences when read on the two strands. (c) A large heterozygous deletion or insertion less than 10 kb in size, with both break-points located within the gene, is evident from heterogeneity in the product sizes in long PCR, with a need to sequence the two alternate products independently to infer the nature and location of the re-arrangement.

A heterozygous deletion or insertion longer than 10 kb or with only one break-point located within the gene does not necessarily manifest in a simple manner by such methods, but can be detected by the use of restriction mapping combined with Southern blotting, or by so-called one-sided PCR methods, in which the flanking regions of a given sequence in genomic DNA are characterized by restriction digestion, dilution, religation to form circles, and PCR using two adjacent, outwardly oriented primers.

Where evidence of a heterozygous mutation is obtained, some form of haplotype analysis is still required to confirm that the CAG repeat-length mutation and the additional heterozygous mutation are on different copies of the gene. This requires a combination of standard mutation detection methods, such as allele-specific PCR, or restriction fragment length polymorphism applied over a considerable length of genomic DNA, involving, for example, long-extension PCR or Southern blotting or a combination of such techniques.

The implications of the results from this much more exhaustive test are similar to those outlined under example 1. Assignment of a mutation may not be straightforward, but *POLG* mutations that cause changes of reading frame, that destroy splice sites or create new ones, that generate a stop codon, that alter a highly conserved amino acid or that delete or interrupt a significant region of coding sequence are unquestionably of pathological significance, if combined with CAG repeat-length mutation in the other copy of the gene.

Detection of a second, clearly pathological mutation in such an individual, outside of the CAG repeat, would confirm that his infertility is of essentially the same kind as those already discussed under example 1, with similar

implications for counselling and treatment.

Example 3

Application of the invention to population-based screening for genetic predisposition to infertility

5 This example describes the way in which the invention is to be applied in the context of genetic screening of the population. It is assumed that this application will respect privacy, plus any relevant legislation in force in different jurisdictions. Its goal is to enable people to make informed choices about their reproductive life without prescribing to them any particular course
10 of action. The possibility for a male to know, in advance, that he will suffer infertility at reproductive age, empowers both him and his partner in a way that other kinds of clinical testing do not, given that they are typically applied long after the problem has manifested and usually too late for its consequences to be completely corrected or managed. This application could be placed in the
15 context of a programme of voluntary genetic testing applied in infancy, or at any age, being offered as one of a large number of tests for genetic disorders. It could also be in the context of a confidential advisory service or even, eventually, a self-test kit. The specific example given is in the context of a voluntary public health programme focused on the specific question of genetic
20 predisposition to infertility and offered in early adolescence.

A defined population of young males (e.g. all those reaching their 17th birthday in a given year in a given locality) are contacted via family doctors, educational institutions or a population registry and asked whether they wish to participate voluntarily in the screening programme under conditions of
25 total confidentiality. The benefits of choosing to do so, plus the predictive limitations of the test, are fully explained to all enabling them to make an informed choice about their participation. Buccal smears (mouthwash samples) are collected from those agreeing to take the test, and DNA is extracted using similar methods as outlined under example 1. *POLG* genotyping is carried out using
30 the same techniques as outlined in examples 1 and 2 above. The results of the test are communicated to all volunteers on a confidential basis, with more specific counselling offered to anyone who requests it.

The test results and their implications are clear: those found to be homozygous for loss of the wild-type *POLG* repeat-length allele, or to be compound heterozygotes with one copy of the gene mutated at the CAG repeat and one copy carrying a clearly pathological mutation elsewhere in the gene,

are advised that prior genetic surveys indicate that they will suffer from a fertility problem, and that if they wish to have children they should consult an infertility clinic at the appropriate time to arrange for assisted reproduction as already described under example 1. Those found to carry at least one wild-type

5 copy of the gene are advised that one common, genetic cause of male infertility has been excluded, but that this does not necessarily mean that they will be free of fertility problems, since there are other genetic and environmental causes that account for a large fraction of fertility problems. Hence they should be aware of the relevant services available in their locality, and advised to

10 seek specific advice if they become aware of a fertility problem in the future.

Obviously, the value of the test described in this example will be greatly enhanced, if combined in the future with tests for other genetic predispositions to male infertility, as these become apparent. This will strengthen the predictive value of the test, although even those in whom all genetic causes of

15 male infertility have been excluded must be made aware that their partner may also suffer fertility problems independently, and that genetics cannot explain everything.